DOCUMENT RESUME

ED 212 266 IR 009 972

AUTHOR Penniman, W. David
TITLE Modeling and Evaluation of On-Line User Behavior. Final Report to the National Library of Medicine.
INSTITUTION OCLC Online Computer Library Center, Inc., Dublin, Ohio.
SPONS AGENCY National Library of Medicine (DHHS/NIH), Bethesda, Md.
PUB DATE Sep 81
GRANT NLM/EMP-(1-R01-LM-03444-01)
NOTE 54p.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Behavior Patterns; *Databases; *Information Retrieval; Information Systems; *Online Systems; Search Strategies; Summative Evaluation; Tables (Data); *User Satisfaction (Information)
IDENTIFIERS National Library of Medicine MD

ABSTRACT

This investigation of search patterns associated with the MEDLINE database had four primary objectives: (1) to refine the stochastic process method of evaluating user behavior; (2) to examine user search behavior exhibited when selecting documents sets; (3) to identify factors that affect the patterns of user behavior; and (4) to use data from the National Library of Medicine (NLM) MEDLINE system to develop a model of ineffective user behavior. Samples of online session protocols obtained from NLM's online system ELHILL-3 were processed in three major phases: the recovery of individual MEDLINE user sessions from the transaction tapes, parsing the individual user and system transactions to assign a unique activity code, and the analysis of user/system behavior using the parsed sessions. The data obtained were analyzed by user ID, daily usage, and level of activity, as well as selected variable analysis and fragment analysis. Findings are outlined and compared with prior studies, and the results of data analyses are summarized for each of seven hypotheses. The methodology and results are detailed, and a need for additional research in the area of search effectiveness is indicated. Twenty references are listed. (RAA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

ED212266

MODELING AND EVALUATION OF ON-LINE USER BEHAVIOR

Final Report to the
National Library of Medicine

Under Extramural Program Grant No.
NLM/EMP (1 RO1 LM 03444-01)

Principal Investigator
W. David Penniman, Ph.D.
Director, Development Division

OCLC

September, 1981
(Revised)

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Lois Yoakam

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR009972

TABLE OF CONTENTS

1.0 Introduction

1.1 Background

1.2 Purpose

1.3 Review of Related Research

2.0 Objectives

2.1 Stochastic Process Method of Evaluation

2.2 User Search Behavior in Document Set Selection

2.3 Factor Discrimination in Patterns of User Behavior

2.4 Approaches to Problem Identification Using MEDLINE Data

3.0 Methodology

3.1 Overview of Methodology

3.2 Hypothesis Testing

4.0 Data Analysis

4.1 Summary Data

4.2 Selected Variable Analysis

4.3 Fragment Analysis

4.4 Comparison of Results to Work by Fenichel and Chapman

5.0 Results

5.1 Refinement of the Stochastic Process Method of Evaluating User Behavior

5.2 Determination of User Search Behavior in Document Set Selection

5.3 Factor Discrimination in Patterns of User Behavior

5.4 Approaches of Problem Identification Using MEDLINE Data

5.5 Additional Results and Findings

6.0 Conclusions

References Cited

ACKNOWLEDGMENT

This study was conducted jointly by the Development Division and the Office of Research of OCLC. Howard Turtle, Senior Research Scientist, provided the software and data reduction for the study with assistance during the early phases by Sande Nissen of the Development Division. Dr. Neal Kaske, Director of Research, served as project manager. Ms. Beckie Purdy provided secretarial support throughout this project and typed the final report.

1.0 INTRODUCTION

1.1 BACKGROUND

Within the past several years techniques have been explored to improve user/system interaction in on-line information retrieval systems by directly monitoring what the user has done during one or more search sessions (CHAPMAN, FENICHEL, MEADOW et al., PENNIMAN, WANGER). These approaches have all drawn upon an approach known as "protocol analysis" in which user transcripts are systematically analyzed in order to learn what people do under certain search circumstances. From such analysis, improvements to search systems can be derived in a variety of ways via:

- Understanding how people actually use a search system (which commands are, or are not used, extent to which features are used in any given search session).
- Comparing search patterns of different classes of users to see if systematic variations can be identified.
- Evaluating where errors are made in the search sequence (e.g. where are most typing errors occurring).
- Creating more informative user messages by incorporating knowledge of what search patterns have occurred thus far and what additional patterns should occur to bring the search to completion.

It is this final area, which has been labelled "adaptive prompting" by Penniman but is also referred to as "intelligent intervention" by RELLES and SONDHEIMER (1981) and "individualized instruction" by MEADOW et al. (1977), that is the focus of the study described in this report.

1.2 PURPOSE

The purpose of the study is to analyze a sample of on-line session protocols obtained from the National Library of Medicine's on-line system ELHILL-3 in order to investigate search patterns associated with the MEDLINE data base. From this investigation, objective conclusions can be drawn regarding the use of the search system and recommendations can be made for improving user/system interaction with an emphasis on the "adaptive prompting" concept.

1.3 REVIEW OF RELATED RESEARCH

Since the early 1970's work has been underway to try to improve user/computer interaction. One example of such work was a project by

W.C. Mann who was concerned with "creating a model of natural human communication which could serve as a guide and a source of corrective suggestions and processes to improve the sorry condition of man-machine communication as it existed (and still exists)" (MANN, 1977). This work involved four basic steps:

capturing human dialogues
identifying important phenomena in the dialogue
creating processes that could follow the dialogue
evaluating the processes relative to the identified phenomena

At the same time, NLM was recording and analyzing human/computer dialogues representing user communication with the MEDLINE system (BROWN and AGRAWLA, 1974). The results from this NLM study, while presented at a general level, gave strong indication of the utility of data gathered via an on-line monitor. While the NLM study addressed only the first two steps identified by Mann, other work was underway which attempted to use monitor or audit-trail data to build tools for improving user interaction.

Penniman presented research results based on monitor data analysis of on-line interaction with an information retrieval system in 1975. This was one of the first studies to analyze monitor data using a stochastic process analysis model for predictive purposes (see also STABELL, 1974). This work has been extended since that time in a variety of areas (e.g. PENNIMAN and PERRY, 1976; PENNIMAN, 1977; DOMINICK and PENNIMAN, 1979; PENNIMAN and DOMINICK, 1979; PENNIMAN and DOMINICK, 1980). The primary goal of all of this work has been to understand and improve the user interface to on-line information retrieval systems and, specifically, to provide more intelligent prompting techniques for naive users. As an initial step, it has been necessary to provide in both a descriptive and a predictive sense the user behavior associated with on-line searching. The transition matrix and related Markov chain was chosen as a mechanism for presenting this data in a condensed, yet meaningful, manner.

In a separate, though similar, effort, ANDREAE (1977) was working with transition matrices to create systems displaying "artificial intelligence." ROUSE (1980) in his work on models of human-machine interaction discussed the relationship between production systems (vis a vis artificial intelligence), pattern recognition and Markov chains for modeling sequences of actions (e.g. in a user-system interaction sense). He suggested that the Markov chain, though not widely used within this context, may have significant potential in future enhancements to the human-computer interface. BURTON and BROWN (1979), on the other hand, argued that the problem of identifying a user's shortcomings via observing his behavior was extremely complex and that, unless the context is very well defined, offering strategies via an automated system may not be very helpful. At the same time,

they also found that for those users provided with an automated "coach" interaction with the system was more enjoyable.

While those concerned with artificial intelligence continue to pursue a variety of avenues involving production systems, etc. there are some information scientists who are extending the work involving transition matrices and associated pattern analysis in hopes that user training and user aids can be enhanced by this mechanism (FENICHEL, 1981;CHAPMAN, 1981). The work by FENICHEL and CHAPMAN built upon previously mentioned work by Penniman. In the current study by OCLC, their work (particularly Chapman's) is in turn built upon in the selection of user/system interaction coding schemes.

While information has been collected at a general level regarding the phenomena of on-line searching across several systems (WANGER, 1976), detailed studies of particular systems and the opportunities for feature enhancement or improved prompting are not extensively available. A recent study done for NLM (WANGER, et al., 1980) represents a significant effort at addressing this area. While that study did not rely upon the unobtrusive audit-trail monitoring capability of the ELHILL system, it did involve the analysis of user-system interactions as reflected in the hardcopy "protocols" of selected experimental search sessions. A number of findings of that study are compared to OCLC's findings in subsequent sections of this report.

2.0 OBJECTIVES

Four primary objectives were established for this research program:

1. To refine the stochastic process method of evaluating user behavior.

2. To examine user search behavior exhibited when selecting document sets.

3. To identify factors that affect the patterns of user behavior.

4. To use data from the National Library of Medicine MEDLINE system to develop a model of ineffective user behavior.

These objectives will be briefly described here. The methodology employed to meet these objectives will be described in Section 3.

2.1 STOCHASTIC PROCESS METHOD OF EVALUATION

The success of adaptive prompting methods will depend upon the availability of reliable metrics for evaluating human/computer interaction. Classical system designs evaluate the interaction in terms of the computer's ability to interpret user input in a nearly context-free manner, with little consideration given to the history of interaction. This classical approach is notable for its simplicity and is well-suited to the identification of most syntactic and some semantic errors. While this type of evaluation may be sufficient for prompting experienced system users, it does little to assist inexperienced and infrequent users who have difficulty formulating overall strategies for satisfying their information needs.

The scope, or class, of user interaction problems that can be recognized by a computer-based system can be expanded by including an historical context. An historical context is the entire history of user interaction in the context in which the interaction is being evaluated. This approach is analogous to the approach used in high-level compilers where both the semantic and syntactic correctness of a body of code is checked.

An adaptive prompting system faces two additional complications when using an historical context. First, the analysis of user interaction is dynamic and must be performed in an incomplete context. This is because the complete context is available only at the conclusion of a session. Since a complete analysis at the end of a session cannot help a user during a session, analysis during a session must employ inferential rather than deductive techniques. Second, the human/computer interaction cannot be represented as a formal language.

While the correctness of program text can be determined within a formal language definition, no such formal definition of correctness exists when dealing with human/computer systems, which, by their nature, are open systems.

The first objective of the research program was to refine the stochastic process method of evaluating user behavior. This objective was met by introducing enough formalism into a session to permit effective analysis without destroying the nondeterministic flavor of the underlying behavior. Three implicit hypotheses underly this objective:

1. User on-line search behavior can be represented as semi-formal state strings.

2. User on-line search behavior exhibits regularities that can be detected as patterns in the state strings.

   The types of regularities that can be detected depend heavily upon the state set chosen. This gives rise to the concept of "state resolution," discussed in the next section.

3. State patterns can provide information about the probable effectiveness of specific user behavior.

2.2 USER SEARCH BEHAVIOR IN DOCUMENT SET SELECTION

Concurrent with refining the adaptive prompting method of evaluation, patterns of behavior used to select document sets in the MEDLINE system were identified. The observed patterns provided insight into the techniques users employed in document searches and the frequency with which these techniques were used.

Two research hypotheses underly this objective:

1. Pattern analysis will reveal the techniques used for document search and frequency of occurrence of the techniques.

2. Pattern analysis will reveal secondary factors influencing when and how specific search techniques are used.

2.3 FACTOR DISCRIMINATION IN PATTERNS OF USER BEHAVIOR

Once patterns of user behavior were determined, samples representing different levels of system use were drawn to allow comparison of pattern usage.

Two research hypotheses underly this objective:

1. Measurable variables can be found that are both pattern-dependent and pattern-independent.

2. Pattern history and pattern-dependent system variables correlate highly with independent assessments of search effectiveness.

2.4 APPROACHES OF PROBLEM IDENTIFICATION USING MEDLINE DATA

The fourth objective was not met during the project in that no useful model of effective or ineffective search behavior was established even though individual components of such models were identified. We feel that the results to date are promising and can form the basis of such models. Attempts to meet this objective within the current grant, however, proved unsuccessful, partly due to the lack of associated independent evaluation of the searches. This is an area that is discussed further in Chapter 6.

3.0 METHODOLOGY

This section summarizes the steps performed in the analysis of the MEDLARS transaction data to obtain the results presented in Sections 4 and 5.

Section 3.1 is a general review of the methodology used and details the processing steps necessary to extract information from the raw transaction data.

Section 3.2 focuses on the techniques used to test the individual research hypotheses. It extends the material presented in Section 3.1 by showing how the general techniques were applied to each hypothesis tested during the course of the study.

3.1 OVERVIEW OF METHODOLOGY

The processing of the MEDLINE transactions consisted of three major phases: the recovery of individual MEDLINE user sessions from the transaction tapes (Section 3.1.1), parsing the individual user and system transactions to assign a unique activity code (Section 3.1.2), and the analysis of user/system behavior using the parsed sessions (Section 3.1.3).

3.1.1 RECOVERY OF USER SESSIONS

A sample of the MEDLINE transaction log used during the study is shown in Figure 3.1. The sample tapes represent eight weeks of system activity from the two copies of the ELHILL-3 system in operation at NLM during January and February of 1980. The tapes duplicate the MEDLINE traffic tapes except for user ID's which were systematically encrypted to prevent identification of actual users. The transaction logs are difficult to interpret in their raw form and must be normalized to facilitate further processing. This normalization involves the following transformations:

1) Transaction tapes are sorted by user ID, date and time. The individual records in the transaction log are written in the order in which they are processed by the ELHILL-3 system so that transactions from multiple user sessions are interspersed in the transaction log. The sort step effectively recovers individual user sessions.

2) Separation of user sessions which accessed data bases other than MEDLINE. Since the search strategies employed in MEDLINE searches were expected to differ from those used in other data bases (e.g. CHEMLINE or TOXLINE), the user sessions were split into two separate samples, those which accessed only the MEDLINE data base and those which accessed any data base besides MEDLINE. User sessions which access only the MEDLINE data base account for 65.4% of all transactions. Note that sessions in which users accessed multiple databases were included in the non-MEDLINE sample.

FIGURE 3.1 - Sample MEDLINE Transaction Log

```
[redacted] 0E2 80007 0254526        TS 2 (TI) : LIZARD:
[redacted] 0E2 80007 0254547     ΔPROG:  SS (6) PSTG (52)    SS 7 /C?  USER:
[redacted] 0E2 80007 0255108        TS 2 (KW) : LIZARD:
[redacted] 0E2 80007 0255130     ΔPROG:  SS (7) PSTG (8)    SS 8 /C?  USER:
[redacted] 0E2 80007 0255366 E      FRT KW
[redacted] 0E2 80007 0255367     ΔPROG:     1  KW  - TOXBIB  KW  - Animal  KW  - Blood Sugar
[redacted] 0E2 80007 0255368     ΔKW  - Endothelium ULTRASTRUCTURE  KW  - Freeze Fracturing
[redacted] 0E2 80007 0255369 Q   ΔKW  - Mechanoreceptors *DRUG EFFECTS  KW  - Mitochondria
[redacted] 0E2 80007 0256276 A      N
[redacted] 0E2 80007 0256277     ΔPROG:     SS 8 /C?  USER:
[redacted] 0E2 80007 0256424        6 AND 7
[redacted] 0E2 80007 0256426     ΔPROG:  SS (8) PSTG (8)    SS 9 /C?  USER:
[redacted] 0E2 80007 0256567        6 AND NOT 7
[redacted] 0E2 80007 0256589     ΔPROG:  SS (9) PSTG (44)    SS 10 /C?  USER:
[redacted] 0E2 80007 0257229 E      STOP Y
[redacted] 0E2 80007 0257230     ΔPROG:     GOOD-BYE!
[redacted] 0E2 80007 0301515        /LOGIN[redacted] /        /
[redacted] 0E2 80007 0301518     ΔPLEASE ENTER YOUR PASSWORD.    USER:
[redacted] 0E2 80007 0302046        [redacted]
[redacted] 0E2 80007 0302048     ΔPROG:     HELLO FROM ELHILL AT NLM.
[redacted] 0E2 80007 0302049 %   ΔYOU ARE NOW CONNECTED TO THE MEDLINE FILE.  A STORED ADDR
[redacted] 0E2 80007 0302395 C      PGF
[redacted] 0E2 80007 0302400     ΔPROG:  SS (1) PSTG (862)    SS 2 /C?  USER:
[redacted] 0E2 80007 0302575        RADIOIMMUNOASSAY
[redacted] 0E2 80007 0302577     ΔPROG:  SS (2) PSTG (2988)    SS 3 /C?  USER:
[redacted] 0E2 80007 0303083        1 AND 2
[redacted] 0E2 80007 0303091     ΔPROG:  SS (3) PSTG (43)    SS 4 /C?  USER:
[redacted] 0E2 80007 0303525        TS :KETO: OR (AB) :KETO:
[redacted] 0E2 80007 0303542     ΔPROG:  SS (4) PSTG (5)    SS 5 /C?  USER:
[redacted] 0E2 80007 0304241        FRT COMPRINCLUDE$
[redacted] 0E2 80007 0304242     ΔUSER:
[redacted] 0E2 80007 0304384        FRT COMPR INCLUDE MH
[redacted] 0E2 80007 0304386 ß   ΔPROG:     1  AU  - Ghodgaonkar RB ; Dubin NH ; Blake DA ;
[redacted] 0E2 80007 0304386 I   ΔMH  - Abortion, Induced ; Adult ; Amniotic Fluid/*ANALYSI
[redacted] 0E2 80007 0304390     ΔCONTINUE PRINTING? (YES/NO)    USER:
[redacted] 0E2 80007 0306268 A      Y
[redacted] 0E2 80007 0306270 P   ΔPROG:     5  AU  - Youssefnejadian E ; Bradovcky H ; Johns
[redacted] 0E2 80007 0306271 e   Δ       Pregnancy ; Prostaglandins F/*BLOOD ; Radioimmunoas
[redacted] 0E2 80007 0307077 E      PGE/BL
[redacted] 0E2 80007 0307083     ΔPROG:  SS (5) PSTG (98)    SS 6 /C?  USER:
[redacted] 0E2 80007 0308019        PGE/  X/BL OR EXP D24.185.78 98.924/BL
[redacted] 0E2 80007 0308026     ΔPROG:  SS (6) PSTG (61)    SS 7 /C?  USER:
[redacted] 0E2 80007 0308232 E      6 OR 5
[redacted] 0E2 80007 0308235     ΔPROG:  SS (7) PSTG (152)    SS 8 /C?  USER:
[redacted] 0E2 80007 0308433        7 AND 2
[redacted] 0E2 80007 0308436     ΔPROG:  SS (8) PSTG (18)    SS 9 /C?  USER:
[redacted] 0E2 80007 0309132        PRT TI, MH COMPR$
[redacted] 0E2 80007 0309132     ΔUSER:
[redacted] 0E2 80007 0309262        PRT TI, MH COMPR
[redacted] 0E2 80007 0309264 \   ΔPROG:     1  TI  - Basal level of human platelet prostagla
[redacted] 0E2 80007 0309265 Y   Δ       Platelet Aggregation ; Prostaglandins E/*BLOOD/ISOL
[redacted] 0E2 80007 0309268 "   Δ       Prostaglandins E/*BLOOD ; Radioimmunoassay ; Rats ;
[redacted] 0E2 80007 0311079 A      N
[redacted] 0E2 80007 0311079     ΔPROG:     SS 9 /C?  USER:
[redacted] 0E2 80007 0311221        PRT TI COMPR
[redacted] 0E2 80007 0311223 J   ΔPROG:     1  TI  - Basal level of human platelet prostagla
[redacted] 0E2 80007 0311223 L   ΔTI  - Influence of angiotensin antagonists on renal vascu
[redacted] 0E2 80007 0311229     Δ       preliminary observations on circulating concentrati
[redacted] 0E2 80007 0311547        PRT TI SKIP5 5
```

3) Multiple system responses were coalesced. During long on-line prints the transaction log may contain more than one system response transaction if the print operation involved multiple terminal writes. Since the transaction logs contain only the first 57 characters of the system response rather than the full text of the on-line print, the multiple transaction records were of little value. Only the first transaction record for a print was retained together with the total number of characters printed for the subsumed records.

4) Each transaction was tagged as a logon request, a system response or a user input to simplify later processing. Even with multiple system responses removed the transaction logs still contain user input transactions without a corresponding system response because some commands elicit no logged response (e.g., "TIME", "CAPS", and queries beginning with the character '@') and as a result of abnormal session termination.

During this initial processing step the transaction records were reformatted to eliminate unused information in the record header, convert binary information to printable form and to reserve space for record type and activity codes which were assigned during the next step.

3.1.2 ACTIVITY CODE ASSIGNMENT

Many of the analysis techniques to be described in Section 3.1.3 view the interactive session as a sequence of user and system activities or states. In order to facilitate further analysis, each user and system transaction is assigned an activity code based on the text of the user input or system response and upon the context in which the transaction occurs. Section 3.1.2.1 describes the initial set of activity codes and Section 3.1.2.2 describes the procedures which effect the assignment.

3.1.2.1 ACTIVITY CODE SELECTION

The initial selection of activity codes was based on information contained in the NLM On-line Services Reference Manual and information provided by MEDLARS training staff. The selection of initial activity codes is constrained by a number of factors:

- The text of the user input or system response is limited to 57 characters so that it is not always possible to determine exactly what has occurred on a given transaction. Multiple line system messages, on-line prints and long search requests are examples of transactions which are often incomplete.

- The selected state set should be detailed enough to allow further analysis to be conducted using only the assigned

activity code; it should not be necessary to re-examine the text of the transaction record. Note that this is an important feature from a privacy standpoint as well.

- The state set must be exhaustive, and the individual states must be mutually exclusive, and it must be possible to assign the state codes under machine control.

The initial state set selected is shown in Figure 3.2 and consists of 42 state codes in 13 major groups. Since 20 of these state codes can be assigned to both user inputs and system responses, the initial state set consists of a total of 62 recognizable activities. The initial state set was designed to preserve as much information as possible about the user sessions. As such, it exhibits more detail than will typically be useful in later analysis. The resolution of this initial set was later adjusted for analysis by regrouping the initial state set into higher level groups.

While most of the state descriptions provided in Figure 3.2 are straightforward, the activity code assignment for user search terms requires some explanation.

The ELHILL-3 system supports a wide variety of search terms and allows the various search term types to be mixed in Boolean requests. Attempts to categorize the various search types when they occur in a multiple term Boolean expression is both difficult, because of the large number of possible combinations, and unreliable, since many such expressions are not completed within the 57 characters captured in the traffic tapes.

In order to eliminate ambiguities in the categorization of search request, the hierarchy of request types shown in Figure 3.2 was developed. A given search request is assigned to the lowest numbered category possible. As an example, variations on the search term FISH would be categorized as follows:

Variation	Activity Code
FISH	A7-single term
FISH:	A6-stem term
ALL FISH:	A3-stem-all
ALL FISH: OR BIRD	A2-Boolean

FIGURE 3.2 - Initial State Set

	Activity Category	Applies to User Input	Applies to System Response	Activity Code	Activity Description
A	Search Terms	x		A1	String or sentence searches
		x		A2	Boolean searches (query)
		x		A3	Stem terms (ALL)
		x		A4	Exploded terms
		x		A5	Range terms
		x		A6	Stem terms (with :)
		x		A7	Single terms
B	Search Element Modification	x	x	B1	Elements
		x	x	B2	Subheadings
		x	x	B3	Other (FILE, LIMIT=)
C	Search Related Commands	x	x	C1	Search structure
		x	x	C2	Index structure
		x	x	C3	Mesh structure (TREE + MESHNO)
		x	x	C4	Reorganizing Search Stack
D	Search Responses		x	D1	PSTG
			x	D2	NP or PSTG (Ø)
			x	D3	MM
			x	D4	*NONE
			x	D8	Multiple response
E	Offsearch	x	x	E1	OFFSEARCH
		x	x	E2	End of OFFSEARCH
F	Storesearch	x	x	F1	Start
		x	x	F2	End
G	Prints	x	x	G1	Print(s)
		x	x	G2	Print Offline (also OFFSEARCH print)
		x	x	G3	Modifying print
H	System Commands	x	x	H1	System information (FILES, NEWS, USERS, TIME, DUMP)
		x	x	H2	Information (HELP and EXPLAIN)
		x	x	H3	Message to system (COMMENT)
		x	x	H4	System language manipulation (CAPS, RENAME, VERSION)

FIGURE 3.2
(Continued)

Activity Category	Applies to User Input	Applies to System Response	Activity Code	Activity Description
I Diagnostics		x	I1	Related to search activities (KEBTRM OVFL, 4 more)
		x	I2	Related to PRINT (LIMIT of 300 DOCUMENTS...)
		x	I3	Related to commands (EXPLANATION, REQUESTED HAS NOT...)
J System Prompts		x	J1	To indicate readiness (SS1/C?, HELLO FROM ELHILL AT NLM)
		x	J2	For decisions (DONE?; SPECIFY #, ALL, OR NONE)
		x	J3	For line input (CONTINUE SUBS, PLEASE ENTER PASSWORD)
K Answers to Questions	x		K1	Defined word responses (YES, NO, UP, DOWN, ALL,...)
	x		K2	Number
	x		K3	Character string
L Session Termination	x	x	L1	Normal (STOP, DONE, STOP Y)
	x	x	L2	Abnormal (ELHILL3 IS ABNORMALLY TERMINATING...)
M Erasures	x		M1	with $, @, no input

3.1.2.2 ACTIVITY CODE ASSIGNMENT

Once the initial set of activity codes was developed, a set of procedures was written to assign activity codes to each of the 2.1 million transactions in the MEDLINE session histories.

The procedures for assigning the activity codes were driven from an external description of the possible transaction text together with information about which activities affect later assignments and which activities can be affected by previous assignments. The use of an externally maintained syntax description simplified modifications to the parsing strategy required to handle undocumented commands encountered in early processing.

As each transaction tape was parsed, sessions which could not be handled automatically were marked for manual review. These unparsable sessions were then coded by project staff who would update the syntax description if possible. The most common source of parsing problems arose from sessions which were terminated by a loss of communications.

3.1.3 ACTIVITY ANALYSIS

A number of analysis techniques have been applied to the parsed MEDLINE sessions in order to identify patterns of system use and to detect differences in the way users with different levels of experience use the system.

The basic approach involved 1) a review of the entire sample to gather usage statistics for each user (Section 3.1.3.1), 2) the selection of three samples representing users with low, moderate, and high usage levels (Section 3.1.3.2), 3) the analysis of usage patterns for each sample (Sections 3.1.3.3 - 3.1.3.6), and 4) the comparison of system usage characteristics between and within samples (Sections 3.1.3.7 - 3.1.3.8).

3.1.3.1 USAGE SUMMARIES

The entire sample was reviewed to obtain the following usage statistics for each user ID:

- Number of sessions completed during the sample period.
- Number of transactions completed during the sample period.
- Total time spent using the system during the sample period.
- Number of days for which the user had at least one session.
- Average number of transactions per session.
- Average session duration.
- Average number of transactions per minute.

During the collection of these statistics, user sessions with fewer than 4 transactions were ignored to eliminate sessions which represent

unsuccessful attempts to logon. The results for the 1519 "active" user ID's are shown in Figure 3.3.

FIGURE 3.3 - Usage Characteristics
(For 1519 "active" users)

	Total	Average Per user
Sessions	39,330	25.9
Transactions	2,104,977	1385.8
Time (HRS:MIN)	7274:12	4:47
Transactions/Session	-	53.7
Minutes/Session	-	11.06
Transactions/Minute	-	4.8

3.1.3.2 SAMPLE SELECTION

While some information about the users in the sample was available (institution type, region, begin date, and type of administrative support), it was not possible to determine levels of experience or training for individual users. Since secondary indicators were not available the observed frequency and duration of system use was taken to be an indicator of user experience, that is, users who consistently made heavy use of the system throughout the eight week sampling period were assumed to be experienced users, while those who made infrequent use of the system were assumed to be relatively inexperienced. See FENICHEL (1981) for a similar classification and analysis of users.

Three sub-samples were drawn which contained all sessions for users whose usage characteristics fell within a selected range. The selection criteria and usage characteristics for the three samples are shown in Figure 3.4. Note that the transaction counts used in Figure 3.4 represent transaction pairs (user input followed by system response) rather than the raw transaction counts shown in Figure 3.3.

FIGURE 3.4 : SUB-SAMPLE CHARACTERISTICS

Sample	Selection Criteria	Characteristics (Parenthesized figures are percentages of total sample)							
		Users	Sessions	Transaction Pairs	Time	Transaction Pairs/Session	Transaction Pairs/Minute	Sessions Per User	Time Per Session
Frequent	20.0 to 23.0 hours of use on at least 25 different days	14 (.92%)	1306 (3.3%)	41208 (3.9%)	296:11:01 (4.1%)	31.6	2.31	93.3	13:37
Moderate	6.0 to 8.0 hours of use on at least 8 but not more than 20 days	46 (3.03%)	1223 (3.1%)	40333 (3.8%)	313:59:20 (4.3%)	33.0	2.14	26.6	15:24
Infrequent	0.5 to 4.0 hours of use on at least 4 different days. At least 5 sessions and not more chan 9 sessions.	149 (9.81%)	1044 (2.7%)	25464 (2.4%)	210:46:01 (2.9%)	24.4	2.01	7.0	12:07
		209 (13.8%)	3573 (9.1%)	107005 (10.2%)	820:56:22 (11.3%)				

3.1.3.3 STATE MAPPINGS

The activity or state codes described in Section 3.1.2.1 are designed to preserve as much detail as possible to allow greater flexibility in subsequent analysis. All subsequent analysis involves the declaration of a "state map" which specifies which initial states or state combinations should be viewed as equivalent. The mapping function is invoked for two basic reasons:

1) It is necessary to control the level of abstraction by viewing multiple states as similar or equivalent for a given type of analysis. The experimenter is free to define a state map which will allow comparison with earlier work (such as FENICHEL and CHAPMANs) or to experiment with new state groupings.

2) It is generally more useful to view states as consisting of transaction pairs. In the bulk of the analysis, a single state consists of a user input followed by a system response. This choice of order (user-system rather than system-user) reflects a tendency to view the ELHILL-3 system as predominantly command driven. The reverse ordering would be more appropriate in menu based systems in which the user responds to a fixed set of options offered by the system. It is worth noting that the ELHILL-3 system, while predominantly command driven, does have a menu driven component (for example, choosing from multiple meanings or answering questions). The user-system ordering was chosen after analysis with small session samples indicated that the system-user ordering generated more than twice as many unique transaction pairs as the user-system ordering. The ELHILL-3 system responds more predictably to user input than users respond to ELHILL-3's prompts.

Two state mappings were used during the analysis. The first mapping was developed to allow comparison with earlier work by Chapman and uses the states shown in Figure 3.5. The second mapping (Figure 3.6) was developed to improve discrimination for different types of search requests and distinguish between successful and unsuccessful searches.

FIGURE 3.5 - Activity Code Mapping Based on User Input

	Activity Code	Description
-	NULL	Transactions which are ignored for analysis purposes
0	STORESEARCH	Storesearch requests
1	ERROR	Erasures of input command or no input
2	NEUTRAL	Commands which do not directly affect search structure (system information and commands)
3	BEGIN	Logon requests
4	DICTIONARY	Requests for information about index or mesh structures
5	TERM	Simple search terms
6	ADVANCED TERM	All searches which are not simple term or boolean and modifications to elements or subheadings
7	BOOLEAN	Boolean searches
8	DISPLAY	On-line print requests
9	END	Requests to end a session
10	OFF SEARCH	Requests for an off-line search
11	PRINT OFF-LINE	Requests for an off-line print

FIGURE 3.6 - Activity Code Mapping Based on User Input and System Response

Activity Name	Description
SRCH_TERMP	Simple term searches which returned postings
SRCH_BOOLP	Boolean searches which returned postings
SRCH_STEMP	Stem searches which returned postings
SRCH_STRGP	String searches which returned postings
SRCH_EXPLP	Exploded searches which returned postings
SRCH_TERMNP	Simple term searches which returned no postings
SRCH_BOOLNP	Boolean searches which returned no postings
SRCH_STEMNP	Stem searches which returned no postings
SRCH_STRGNP	String searches which returned no postings
SRCH_EXPLNP	Exploded searches which returned no postings
SRCH_TERMMD	Simple term searches which resulted in a request for more information
SRCH_BOOLMD	Boolean searches which resulted in a request for more information
SRCH_STEMMD	Stem searches which resulted in a request for more information
SRCH_STRGMD	String searches which resulted in a request for more information
SRCH_EXPLMD	Exploded searches which resulted in a request for more information
SRCH_ANSWP	Search answer resulting in postings
SRCH_ANSWNP	Search answer resulting in no postings
SRCH_ANSWMD	Search answer resulting in a request for more information
OFFSRCH	Offsearch request
BAD_ANSWER	Incorrect answer to a system request for information
PRNT	Print request
PRNT_BAD	Incorrect print request
PRNT_OFFLN	Offline print request
PRNT_MRINF	Print request resulting in a request for more information
ERASURE	User erasure of input line or no input
SYS_INFO	Request for information about the system
LOGON	Logon request
SRCH_MISC	Miscellaneous commands affecting searches (e.g. restack)
LOGOFF	Request to end a session
STORESRCH	Store search request
MISC_ANSWR	Miscellaneous answer to a system question

3.1.3.4 SESSION ANALYSIS

The three samples contain a record of all activity conducted by selected users during the sample period. These samples are viewed as a sequence of user sessions in which each session is a sequence of activities. Each activity is defined in the state map and corresponds to a user input/ system response pair in the original transaction files. A dummy end state is automatically added to each session to simplify processing; all user sessions begin with a logon request and end with the dummy end state. A predefined null state may optionally be used for activities which are of no interest for a given state decomposition.

During session analysis we were concerned with collecting basic session and state statistics, first-order transition probabilities and with the recognition of common state sequences and cycles. In order to obtain this information, the session histories were used to build a weighted directed graph (digraph) in which each node represents a defined state and each edge represents a transition to a new state. The digraph representation is equivalent to the transition matrices ordinarily used for Markov analysis, but has computational advantages for models with a large number of states and when searching for common state sequences. A review of the application of digraphs to the study of Markov processes may be found in DEO (1974); GONZALEZ and THOMPSON (1978) review the use of digraphs in syntactic pattern recognition.

3.1.3.5 SIGNIFICANCE TESTS

In order to test the significance of the different patterns of use detected in the three samples, the Kolmogorov-Smirnov test was applied to the observed frequency distributions for zero, first, and second order transition patterns. Since the Kolmogorov-Smirnov test is sensitive to cumulative differences in frequency distribution, the states were reordered by composite frequency to produce the most conservative test possible.

The results of the between sample tests showed that the sample differences were significant at the .001 level for all three comparisons (bottom to top, bottom to middle, middle to top) and for all frequency distributions (zero, first and second order transitions).

Each sample was then split into two subsamples and the Kolmogorov-Smirnov test applied to detect significant differences within each sample. No differences were detected with significance thresholds set at .05, .01 and .001.

3.2 HYPOTHESIS TESTING

The current research program was designed to test 10 research hypotheses organized under 4 primary objectives. This section reviews the steps taken to test these objectives. Specific results of the tests are presented in Sections 4 and 5.

Objective: The refinement of the stochastic process method of evaluating on-line user behavior.

Hypotheses	Methodology
1. User on-line search behavior can be represented as semi-formal state strings.	The individual sessions in the traffic tapes were recovered and normalized. A set of mutually exclusive activities or states was defined and each MEDLINE transaction was assigned to its corresponding activity, resulting in a representation of each user session as a sequence of state codes.
2. User on-line search behavior exhibits regularities that can be detected as patterns in the state strings.	Frequency distributions describing the occurrence of individual states, state pairs and state triplets were obtained. The observed frequency distributions of state pairs and triplets were compared to predicted frequency distributions based on lower order observations. A user/system ordering was selected for further analysis since it resulted in fewer unique pairs than the system/user ordering.
3. State patterns can provide information about the probable effectiveness of specific user behavior.	State sequences leading to null sets, corrected errors and print statements were obtained. Descriptive statistics were obtained for session length (time and transactions) and for transaction rates.

Objective: Determine user search patterns in selecting document sets.

1. Pattern analysis will reveal the techniques used for document search and frequency of occurrence of the techniques.	Zero, first, and second order transition matrices were generated. Digraph representations were used to search for high probability transition sequences.
2. Pattern analysis will reveal secondary factors influencing when and how specific search techniques are used.	Three session samples were selected to represent light, moderate, and heavy system use. Transition matrices were generated and compared for each sample. Significance tests were run comparing zero, first and second order frequency distributions for the three samples. Significance tests were applied to detect within sample differences.

Objective: Identify factors that discriminate patterns of user behavior.

1. Measurable variables will be found that are both pattern-dependent and pattern-independent.	Differences in the zero, first, and second order frequency distributions for the three samples were isolated. Between sample differences for basic descriptive statistics (e.g., time per session, transaction rate, and session length) were examined.
2. Pattern history and pattern-dependent system variables correlate highly with independent assessments of search effectiveness.	Not tested.

Objective: Develop and test a model of ineffective search behavior.

1. Patterns and associated variables can be used to form a step-like computable probability function that reflects search ineffectiveness.	Not tested.
2. Historical context represented by state strings can be used to improve problem identification specificity once the preset probability threshold has been exceeded.	Not tested.
3. The increased specificity of problem identification can be used to improve the assistance that the system can offer to the interactive user.	Not tested.

4.0 DATA ANALYSIS

The data collected for the sample sessions were analyzed in a variety of ways. Initially, it was useful to characterize the sample using basic summary statistics. Following this, detailed analyses of particular subsamples and activity sequences were evaluated.

4.1 SUMMARY DATA

4.1.1 BY USER ID

While there were 2287 authorized user ID's for the period sampled, only 1523 (or 67%) appeared in the sample. It should also be noted that of the 4905 unique ID's observed in the sample, 3382 (or 69%) were spurious and generated by users attempting to logon with an incorrect (or mistyped) ID. This would indicate that ID entry is still a problem for some users.

4.1.2 BY DAILY USAGE

While the sample covered 43 days in which usage occurred, a total of 39 week days were included. The week-end days had significantly lower usage (by a factor ranging from 10 to as much as 100 in number of users and number of sessions). Elimination of these week-end days from the summary data provides a more realistic indication of level of usage of the system in sessions devoted exclusively to searching of the MEDLINE data base. Tables 4.1 and 4.2 summarize the daily statistics (in terms of totals and averages for number of unique ID's, number of sessions, number of transactions, connect time, transactions per session, connect time per session, and transactions per minute) for the total sample and for week days only.

4.1.2 BY LEVEL OF ACTIVITY

It was observed that many sessions had very low levels of activity (often with no more than two transaction pairs to a session). This led to a further screening of the sample to eliminate "inactive" users (i.e. with transactions below a set threshold). This screening eliminated all the spurious ID sessions because they never got past logon. In addition it eliminated four of the 1523 authorized ID's and their associated sessions since they didn't get past the logon process either. The sessions associated with the remaining 1519 "active" user ID's were then evaluated in terms of daily statistics. Table 4.3 summarizes the results of this evaluation and indicates the impact of dropping almost 12% (or over 5000) of the sessions from the sample where these sessions were essentially aborted efforts to logon. The remaining sessions accounted for 99.5% of the transactions and 99.96% of the connect time.

TABLE 4.1 - Summary Data for Daily Use Statistics
(Total Sample)

	Number of Unique ID's	Number of Sessions	Number of Transactions	Total Connect Time (HR:MIN)	Transactions Per Session	Time Per Session (MIN:SEC)	Transactions Per Minute
Total (43 Days)	1523	44,532	2,115,381	7276:43	-	-	-
Average Per Day	557	1,036	49,195	169:14	47.5	9:48	4.8

TABLE 4.2 - Summary Data for Daily Use Statistics
(Week Days Only)

	Number of Unique ID's	Number of Sessions	Number of Transactions	Total Connect Time (HR:MIN)	Transactions Per Session	Time Per Session (MIN:SEC)	Transactions Per Minute
Total (39 Days)	1523	44,374	2,110,618	7257:20	-	-	-
Average Per Day	612	1,138	54,118	186:05	49	9:59	4.9

TABLE 4.3 - Summary Data for Daily Use Statistics
(Active Users Only)

	Number of Unique ID's	Number of Sessions	Number of Transactions	Total Connect Time (HR:MIN)	Transactions Per Session	Time Per Session (MIN:SEC)	Transactions Per Minute
Total (43 Days)	1519	39,330	2,104,977	7274:12	-	-	-
Average Per Day	450	915	48,953	169:10	53.7	11:06	4.8

In looking at the full protocol for selected sessions having low levels of activity it was observed that sessions having only four or six transaction pairs are also likely problem sessions. In these cases the user is again probably having trouble getting past the logon process (forgotten password, miskeyed ID, etc.). This led to a further screening of the sample to not only weed out sessions with no productive interactions, but to categorize subsets of sessions (and ID's) by the level of activity observed. Prior to arriving at the three subsets identified in Figure 3.4 a number of screenings on usage by user ID were tried including:

- by number of days
- by number of sessions
- by number of transactions
- by total connect time
- by transactions per session
- by time per session
- transactions per minute

The user ID's and associated statistics resulting from these screens led to an analysis of the "top" and "bottom" 20 user ID's. Note that the "bottom" screen required minimum cut off points as listed below:

Category	Minimum Cut-Off
Days of Medline Use	5 days
Total Medline Sessions	5 sessions
Total Number of Transactions	100 transactions
Total Elapsed Time	1 hour
Transactions per session	20 transactions/session
Elapsed time per session	10 minutes/session
Transactions per minute	10 transactions/minute

These initial screenings were useful because they represented an early stage of pattern and problem identification. As an example, a spurious user ID that showed up in the "top 20" list using both the "day" and "session" screen was /LOGI indicating that users are in need of additional instruction on how to respond to the LOGIN command and when to enter the user ID. This error occurred in 39 of the 43 days.

In looking at the user ID's that appeared in the top and bottom lists on selected level of usage screening criteria and matching this with institution start dates (for day and session screens only), the top users had consistently earlier institutional start dates (1974 average start date for top users as compared to 1976/1977 start date for bottom users). The implications of this phenomenon can best be drawn by NLM itself by investigating the change in training patterns, institutional recruiting and support, and the specific environments of the institutions selected by these screening criteria. Specific data on the encrypted user ID's can be provided upon request.

The "top" and "bottom" users in the screenings by number of days active and by number of sessions reflected a pattern as to institutional type as well. For the top 20 users in both screening sets over half came from medical schools or society, foundations or research institutions. The bottom 20 users displayed no such pattern with respect to institutional setting.

Because the institutional ID's were encrypted we could only distinguish between (but not identify) institutions. This early screening of the "top" and "bottom" ID's led to the subsequent screening identified in Figure 3.4 as mentioned earlier.

4.2 SELECTED VARIABLE ANALYSIS

In addition to general summary data, several variables of special interest were analyzed individually to answer questions such as: "Where are errors most frequently made?" or "What search approaches are most likely to result in zero-set results?" The following sections describe several of these specific analyses.

4.2.1 ERROR ANALYSIS

While it is very difficult to identify subtle errors (which was one of the intentions of this study) it is relatively simple to identify in the machine readable protocols those actions where the user introduced a correction. By mapping selected user inputs into a category specified as "erasures" (see Figure 3.2) it is possible to deduce what the user was attempting to do by evaluating the action immediately following the erasure. Figure 4.4 summarizes by percentage where errors are made most frequently that are caught and corrected by the user. Note that over 70% of all such entry errors involve only four basic activities (i.e. Boolean logic, single terms, record display, string search). Note also that over 7% of all erasures in the sample were followed by another erasure (errors upon errors). This data is based on a sampling of over 1300 errors taken from one transaction log tape.

FIGURE 4.4 - Analysis of Where[1] Entry Errors Are Made (and Corrected) by Users (Single Tape Sample)

Activity Code	Activity Category	Frequency	%	Cumulative %
A2	Boolean Searches	475	35.5	35.5
A7	Single Terms	253	18.9	54.4
G1	Print	113	8.4	62.8
A1	String Searches	101	7.5	70.3
M1	Erasures	101	7.5	77.8
A4	Exploded Terms	74	5.5	83.3
C4	Reorganizing Search Stack	39	2.9	86.2
A3	Stem Terms (ALL)	29	2.2	88.4
B2	Search Element Modification	29	2.2	90.6
E1	Offsearch	25	1.9	92.5
L1	Session Termination	25	1.9	94.4
A6	Stem Terms (with:)	22	1.6	96.0
K1	Defined Question Response	21	1.6	97.6
G2	Print Offline	17	1.3	98.9
C2	Index Structure	8	.6	99.5
C1	Search Structure	2	.1	99.6
H1	System Information	2	.1	99.7
E2	End of Offsearch	2	.1	99.8
C3	Mesh Structure	1	.1	99.9

(1) Based on the transaction immediately following an erasure.

4.2.2 SEARCHES WITH ZERO-SET RESULTS

Again, using the activity categories of Figure 3.2 it is an easy matter to evaluate those sequences of activities that resulted in zero set or "no-postings" results. Figure 4.5 presents sample data taken from one transaction log tape. Two-thirds of all such occurrences result from the execution of a range term search in this sample. Over 75% of all such occurrences can be accounted for by two types of activities (range term searching and Boolean searches).

A more detailed analysis of the sequence of activities leading up to the no-posting results (looking back six to eight transactions) indicated that users tend to repeat the same activity in a "looping" fashion despite unsuccessful results. As an example a frequently occurring pattern was repeated entry of a single search term. This sequence represents an attempt to find data base vocabulary that matches the searcher's vocabulary and could be followed by a prompt suggesting alternative means of finding terms with postings. As expected, a second highly repetitious pattern was the continued request to print more documents until a set was exhausted. This sequence is actually an expected use and a diagnostic prompt would not be necessary.

Of more interest is the large number of unique patterns generated leading up to a "no postings" result. For example, when looking back eight steps over 1400 individual patterns were generated using a one-tape sample and no single pattern accounted for more than about 2% of the occurrence of a "no postings" result. If the sequence was shortened to six transactions no single sequence accounted for more than 6% of the occurrences. Consequently, it would seem that the analysis presented in Figure 4.5 would suffice to draw conclusions regarding the "no posting" response and the activities leading to such a response unless a more general mapping of steps can be derived (e.g. possibly the string concept described by Chapman).

FIGURE 4.5 - User Activities Resulting in "No Postings" Response (Single Tape Sample)

Activity Code	Activity Category	Frequency	%	Cumulative %
A7	Single Terms	1605	67.2	67.2
A2	Boolean Searches	244	10.2	77.4
A4	Exploded Terms	160	6.7	84.1
A6	Stem Terms (with:)	149	6.2	90.3
A3	Stem Terms (ALL)	98	4.1	94.4
K1	Defined Response	88	3.7	98.1
A1	String Search	22	.9	99.0
G1	Prints	11	.5	99.5
K2	Number Answer	4	.2	99.7
C3	Mesh Structure	2	.1	99.8
L1	Normal Termination	2	.1	99.9
G3	Modify Print	1	-	
M1	Erasure	1	-	
C1	Search Structure	1	-	
C4	Reorganize Stack	1	-	
E1	Offsearch	1	-	

4.2.3 PATTERNS SUGGESTED BY PREVIOUS NLM RESEARCH

A study by WANGER, et al. (1980) suggested that users may exhibit at least two types of "non-productive" behavior:

hopscotching - jumping from approach to approach without evaluating any one in detail

thrashing - staying with one approach and trying all combinations without looking at documents, but only at set size

These patterns are similar to those suggested by MEADOW and built into his IIDA user diagnosis software, but have not been verified empirically as actually representing real user behavior.

Consequently, we decided to analyze selected data from our sample in such a way as to evaluate the premise in the WANGER report that users didn't browse documents in refining their searches (i.e. a form of "thrashing"). In performing this analysis two techniques were used. First, the state mappings shown in Figure 3.5 were evaluated for the three samples described in Figure 3.4 (i.e. frequent, moderate, and infrequent users). The results of these analyses indicated that the display command was invoked an average of 3.0 (infrequent users), 4.3 (moderate users), and 3.4 (frequent users) times per session. Further, the display command was followed by another search related activity (term, advanced term, Boolean) 45% of the time by infrequent users, 52% by moderate users, and 52% by frequent users. So it would appear that the display command is part of a continuing search sequence in about half of the cases where it is invoked. This is substantiated by looking at the most frequently occurring sequence of activities where a "display" activity occurs in the sequence at the beginning and at the end (i.e. a loop of activity length (n) around the display command). Over 40% of all such sequences (for frequent users) were of a pattern indicating search refinement. The remainder of such loops were repeated display commands indicating that the user was most likely stepping through a lengthy set.

As a result of these analyses it would appear that all three classes of users do, in fact, use the display command to browse or refine their searches at least part of the time.

Other areas suggested by WANGER, et al. for further research included: (1) the use of the stringsearch capability and (2) the use of the truncation feature. Data were not analyzed regarding these two areas, but are available for analysis should NLM wish to do so in a subsequent study.

4.3 FRAGMENT ANALYSIS

By using the activity mappings shown in Figure 3.5 sequences of activities of selected lengths can be analyzed for each of the three user categories (infrequent, moderate, and frequent). For the analyses in this study fragments of length 2, 3, and 4 were chosen. Fragment length in this sense means the number of activities that were looked at in sequence. As a starter it is interesting to note the number and distribution of sequences by length and user type. For users sampled in the three categories 80% of all fragment occurrences were accounted for by anywhere from 13% to 22% of the fragment types occurring. This means that of all sequences of events that did occur, a relatively small number of sequences accounted for most of the occurrences (a real example of the 80/20 rule). The actual percentages are shown in Figure 4.6.

FIGURE 4.6 - Percentage of Fragments of Lengths 2, 3, and 4 Accounting for at Least 80% of All Activity by User Category

User Category	Percentage of Fragments Accounting for 80% of All Occurrences by Fragment Length		
	2	3	4
Infrequent	18%	15%	22%
Moderate	18%	14%	19%
Frequent	17%	13%	17%

The analysis of fragments can provide insight into not only the most frequently occurring sequences, but also the diversity of sequences. This is particularly true as longer sequences are evaluated. For example, the number of different types of sequences of length 4 that occurred were as follows:

infrequent users:	1886 fragments
moderate users:	2096 fragments
frequent users:	2017 fragments

The number of unique fragments accounting individually for at least 10% of the occurrences were as follows:

infrequent users:	9 fragments
moderate users:	9 fragments
frequent users:	10 fragments

This is strong evidence that the level of specificity of the activity codes can not be increased (e.g. to our earlier set of codes shown in Figure 3.2) without generating an even more diverse set of fragments. The implications of this for being able to identify expected sequences of activities by merely counting occurrences are dire. This should not imply, however, that a detailed analysis of more specific activity codes in certain areas is not warranted. In fact, in areas such as search term entry, Boolean structures, and stringsearches (to name a few) more detailed activity codes would be extremely useful in analysis.

If only the top ten fragments of length 4 are selected for each of the three user samples, we find a 60% overlap between all three samples (see Figure 4.7). That is, six of the ten most frequently occurring fragments in each sample occur in the top ten fragments of each sample. This is further indication that longer fragments not only lead to a larger number of fragments (as expected), but that longer fragments may not help in distinguishing between users since the most frequently occurring fragments are nearly identical for each group and the remaining fragments do not occur frequently enough to draw any conclusions. This points to the need for reliance upon the analysis of shorter fragment lengths (which do in fact distinguish between each of the three user groups as indicated in Section 3.1.3.5) followed by a more detailed analysis of selected activity sequences.

FIGURE 4.7 - Comparison of the Ten Most Frequently Occurring Fragments of Length Four for Each of Three User Groups

USER GROUP

Frequent	Moderate	Infrequent
7-7-7-7**	7-7-7-7**	7-7-7-7**
7-7-7-8**	7-7-7-8**	8-8-8-8
6-7-7-7**	7-8-7-8*	7-7-7-8**
7-7-7-6	7-7-8-7**	5-5-5-5*
7-7-6-8	7-8-7-7-**	7-7-8-7**
7-7-8-7**	5-7-7-7**	7-8-7-8*
7-8-7-7**	5-5-5-5*	7-8-7-7**
5-7-7-7**	8-7-8-7	5-7-7-7**
8-7-7-7	6-7-7-7**	6-7-7-7**
7-6-8-7	5-5-7-7	5-7-5-7

* Occurs in one other list

** Occurs in two other lists

The importance of the analysis of shorter fragment lengths is thus suggested in that it not only is much easier to evaluate (due to fewer possible fragments) but may actually provide better contrast between users for the design of a set of prompting analyzers.

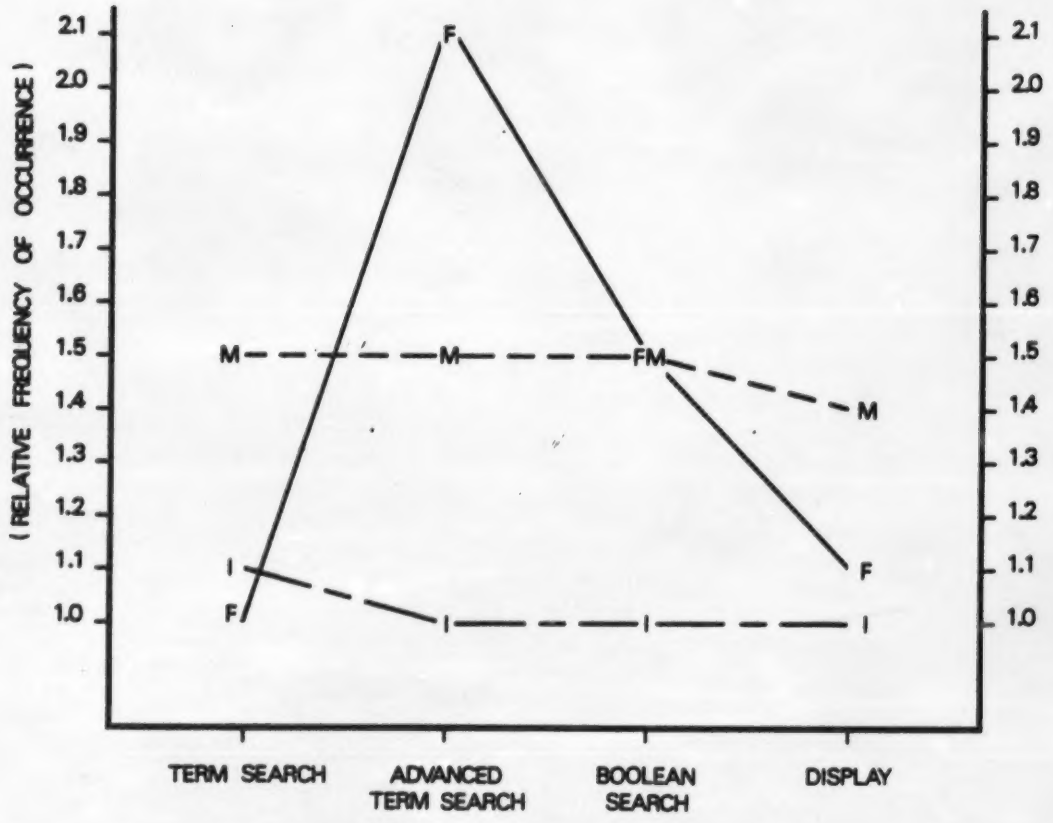
As indicated in Section 3.1.3.5, the Kolmogorov-Smirnov test indicated significant differences between each of the subsamples for fragments of length one, two, and three (zero, first, and second order analyses). The natural question, then, is "What makes them different?" An analysis of each activity provides the profiles illustrated in Figure 4.8. Major differences occur in the frequency with which each group uses a variety of system features. The most significant differences are in those activities which not only are used more by one group but also occurred with some frequency in any of the sample groups. These are the term search, advanced term search, Boolean search and display activities. Other activity differences may appear interesting to pursue (e.g. stored search, dictionary look-up, off-line search) but occurred too infrequently across all three subsamples to draw any conclusions - except that they are features that are used in very few search sessions. For the frequently occurring activities we can draw user profiles as shown in Figure 4.9. This indicates that the frequent (or experienced) searcher is likely to use about the same number of single terms and displays in a search as an infrequent user, but twice as many advanced term search entries and half again as many Boolean searches. The moderate user, on the other hand is likely to use more of everything than the infrequent user and more single terms and displays than the frequent user. This would suggest different prompting strategies for each of these user categories as well as teaching different strategies in beginning and advanced training classes.

Similar analyses were done for first-order data (i.e. fragments of length two) and significant differences were found in those sequences starting with Boolean, single term, and some advanced term sequences when comparing moderate to frequent users. Comparisons between moderate and infrequent users indicated significant differences in sequences involving Boolean, search displays, single term searches, advanced term searches, and the first steps after beginning the session. Detailed data on these comparisons are available in both printed and machine readable form.

FIGURE 4.8 - Comparative Profiles for Frequent/Moderate/Infrequent Users Based Upon Ratios of Frequency of Occurrence of Activities

	Activity Type	Ratio of Frequency of Occurrence By User Category		
		Frequent	Moderate	Infrequent
0	Stored Search	7.0	1.0	1.6
1	Error	1.3	1.5	1.0
2	Neutral	1.2	1.7	1.0
3	Begin	1.0	1.0	1.0
4	Dictionary Look-up	1.0	1.9	1.9
5	Term Search	1.0	1.5	1.1
6	Advanced Term Search	2.1	1.5	1.0
7	Boolean Search	1.5	1.5	1.0
8	Display	1.1	1.4	1.0
9	End Search	1.0	1.0	1.0
10	Off Search	1.6	1.8	1.0
11	Print Off-line	1.1	1.0	1.5

FIGURE 4.9 - Comparative Profiles of
Frequent/Moderate/Infrequent
Users for Selected Activities



4.4 COMPARISON OF RESULTS TO WORK BY FENICHEL AND CHAPMAN

The choice of activity codes at the detailed level (shown in Figure 3.2) is unique to this study. As has been pointed out, however, these activities can be mapped onto other similar activity codes for analysis at a less detailed level. The codes shown in Figure 3.5 are similar in most respects to those used by CHAPMAN (1981) in her analysis of searches on the DIALOG system. In addition, the work by FENICHEL (1981) involving the DIALOG system and the ERIC data base provides some useful findings to be compared to our results. Some of the findings in each of these studies are compared to our findings in the following sections. It should be noted that both of those studies involved the DIALOG system and neither involved medical data base searches.

4.4.1 COMPARISON TO FENICHEL STUDY

FENICHEL (1981) evaluated users with different levels of experience as to their search behavior. Several conclusions were drawn which can be compared to the present study. The following sections provide comparisons for selected findings.

4.4.1.1 NOVICES SEARCH MORE SLOWLY

This conclusion matches our findings in that infrequent searchers were slower by 13% (2.01 transaction pairs per minute versus 2.31 for frequent users) and moderate users fell in between (7% slower than frequent users or 2.14 transaction pairs per minute). See Figure 3.4 for data.

4.4.1.2 NOVICES MAKE MORE ERRORS

This conclusion was not consistent with our findings. In fact, we found that frequent users made 35% more errors than infrequent users and moderate users made 50% more errors than infrequent users. It should be noted that the difference in findings could be in the way in which "errors" are defined. Recall that our definition involved the identification and correction of mistakes many of which were probably typographical while Fenichel concentrated on nontypographical errors.

4.4.1.3 EXPERIENCED SEARCHERS USE MORE COMMANDS AND CONNECT TIME THAN INEXPERIENCED USERS

While our conclusions agree for the frequent and infrequent users (32 transaction pairs versus 24 transaction pairs) Fenichel's moderate user group was lower than either extreme group and ours was higher. (See Figure 3.4). As for connect time, the same was true, i.e. our experienced users took longer than inexperienced users but moderate users took longer than either group (frequent = 13 minutes; moderate = 15 minutes; infrequent = 12 minutes). Again, Fenichel's study showed the moderate group taking less time than either experienced or novice users.

4.4.1.4 LARGE INDIVIDUAL DIFFERENCES EXIST IN SEARCH BEHAVIOR BETWEEN USERS OF SIMILAR EXPERIENCE

This phenomenon was not adequately tested in our study to draw a firm conclusion although the data certainly exists. We did perform a split half analysis of the data for experienced users and found that no significant differences existed between the two data sets in terms of commands used or sequence of commands. It should be noted that further analysis is needed for an adequate comparison.

4.4.1.5 EXPERIENCED USERS EMPLOY RELATIVELY SIMPLE SEARCH STRATEGIES

Fenichel found that for experienced users over half of the searches involved no refinement beyond the initial strategy and it was seldom that any but the most basic techniques for selecting and combining terms were employed. This was not the case in our findings. In fact, the experienced users were twice as likely to use advanced search techniques (see Figure 4.9) and moderately experienced users fell in between.

4.4.2 COMPARISON TO CHAPMAN STUDY

CHAPMAN (1981) conducted several comparisons involving the influence of user training and system diagnostics on user behavior. One specific comparison involved test subjects with different levels of experience (trained, but inexperienced users versus professional searchers). Our findings are compared to the results of her evaluation of those two groups of users as well as to some general findings of her study.

4.4.2.1 INCREASE IN ORDER OF ANALYSIS INCREASES NUMBER OF UNIQUE STATE STRINGS GENERATED

This result agrees with our findings as described in Section 4.3. The implication of this finding is that while higher order analysis increases our ability to predict actions, it does not reduce the complexity of the task.

4.4.2.2 DIFFERENCES IN LOW ORDER DATA APPEAR TO BE SUFFICIENT TO ACCOUNT FOR DIFFERENCES IN TRANSITIONS OF HIGHER ORDER

Our findings do not support this conclusion. In fact, while differences may be identified by using simple frequency counts of command usage, higher order transitions cannot be predicted from this data. Our tests show that actual higher order transition data are significantly different than data generated artificially from lower order transition matrices. Consequently, higher order analysis is necessary to accurately predict longer sequences of activities.

4.4.2.3 EXPERIENCED SEARCHERS USE MORE ADVANCED SEARCH TECHNIQUES WHILE INEXPERIENCED SEARCHERS USE MORE INDEX SEARCH AND DISPLAY COMMANDS

Our findings verify that experienced users employ more advanced search techniques. For inexperienced users we also found that single term search and dictionary look-up were employed more often. We did not find, however, that inexperienced users employed the display function more often. (See Figure 4.8). Our "moderately experienced" users, in fact, employed the display function more frequently and behaved more in tune with Chapman's findings for inexperienced users in this area.

4.4.2.4 COMPARISON OF FREQUENTLY OCCURRING ACTIVITY SEQUENCES

Chapman presents some results that can be compared to our findings shown in Figure 4.7. Her results for experienced and inexperienced users were available in only the briefest form, but indicate that the most frequently occurring sequence of activities of length four for inexperienced users is a sequence of "display" commands. Our results would tend to confirm this in that for our inexperienced users this sequence was the second most frequent sequence (topped only by sequences of Boolean combinations). Our results do not agree with hers for experienced users, however, where she found that advanced select sequences were the most frequently occurring. We found nineteen other sequences occurring more frequently than this one (thus its absence from our Figure 4.7).

5.0 RESULTS

The analysis of data described in Chapter 4.0 in many cases extended beyond the initial hypotheses. In this chapter, we return to those hypotheses and summarize the findings for each. Where results were inconclusive or inadequate, suggestions for additional work are included.

5.1 REFINEMENT OF THE STOCHASTIC PROCESS METHOD OF EVALUATING USER BEHAVIOR

This objective involved three hypotheses which are repeated below along with our conclusions based on the associated research.

5.1.1 HYPOTHESIS: USER ON-LINE SEARCH BEHAVIOR CAN BE REPRESENTED AS SEMI-FORMAL STATE STRINGS

This hypothesis is verified in that sets of activity codes were developed which could accommodate all transactions in the audit trail tapes supplied by NLM. State strings were derived at several levels of specificity depending upon the activity codes selected.

5.1.2 HYPOTHESIS: USER ON-LINE SEARCH BEHAVIOR EXHIBITS REGULARITIES THAT CAN BE DETECTED AS PATTERNS IN THE STATE STRINGS

This hypothesis is verified in that a relatively small set of sequences of activities were derived that accounted for the bulk of interactions. At the same time, however, as string length was increased a large number of relatively unique sequences were generated that could not be condensed even with more general activity codes. So, while regularities do exist they don't account for <u>all</u> forms of interaction.

5.1.3 HYPOTHESIS: STATE PATTERNS CAN PROVIDE INFORMATION ABOUT THE PROBABLE EFFECTIVENESS OF SPECIFIC USER BEHAVIOR

Patterns and frequency data were used to evaluate error, null set and display command occurrence to learn more about each of these individual components of search effectiveness. In addition, patterns, and even raw frequency data for activity occurrence, were also used to distinguish between subsamples drawn from frequent, moderate, and infrequent users. In that sense the hypothesis was verified. That measure of effectiveness (i.e. frequency of use) is relatively weak, however, and more research is needed on how to deduce "effectiveness" from available information in audit trail data. One suggested approach is to evaluate the set sizes associated with terminating displays and whether or not searches are terminated with a display activity.

5.2 DETERMINATION OF USER SEARCH BEHAVIOR IN DOCUMENT SET SELECTION

This objective involved two hypotheses:

5.2.1 HYPOTHESIS: PATTERN ANALYSIS WILL REVEAL THE TECHNIQUES USED FOR DOCUMENT SEARCH AND FREQUENCY OF OCCURRENCE OF THE TECHNIQUES

Frequency counts of activity code sequences did reveal the commonly used approaches to document set selection. As indicated in the findings for Hypothesis 5.1.2 however, there were a large number of sequences generated that had relatively small frequencies - even with use of a more general set of activity codes. Therefore, more work needs to be done in the area of pattern analysis. String sequences such as employed by CHAPMAN may help in this area.

5.2.2 HYPOTHESIS: PATTERN ANALYSIS WILL REVEAL SECONDARY FACTORS INFLUENCING WHEN AND HOW SPECIFIC SEARCH TECHNIQUES ARE USED

Both the similarities and differences between users with different levels of experience were evaluated. Connect time, rate of interaction and use of certain commands were found to be distinguishing characteristics. Significant differences were found not only in the frequency with which various commands were employed, but also in the sequence of commands employed.

5.3 FACTOR DISCRIMINATION IN PATTERNS OF USER BEHAVIOR

Samples were drawn from the total sample based on a preliminary analysis. These subsamples as indicated previously represented groupings of sessions by user ID based on frequency of use of the system and provided three subsamples representing frequent, moderate and infrequent users.

5.3.1 HYPOTHESIS: MEASURABLE VARIABLES WILL BE FOUND THAT ARE BOTH PATTERN DEPENDENT AND PATTERN INDEPENDENT

The findings do indicate that for the variable used as a selection criterion (i.e. frequency of use) there were both similarities and differences in pattern occurrences.

5.3.2 HYPOTHESIS: PATTERN HISTORY AND PATTERN-DEPENDENT SYSTEM VARIABLES CORRELATE HIGHLY WITH INDEPENDENT ASSESSMENTS OF SEARCH EFFECTIVENESS

This hypothesis was not tested in that no independent assessments of the searches were developed. The assessment of search effectiveness by other than traditional means (i.e. recall and

precision) is an area needing further research. While we were not able to develop such assessments in this study, we still believe it is possible given the tools we have now developed and the pool of data across different systems (e.g. work of Fenichel and Chapman) that can now be compared.

5.4 APPROACHES OF PROBLEM IDENTIFICATION USING MEDLINE DATA

While the overall objective in this area was not met (in that no useful model of effective and ineffective behavior was derived) several individual areas of problem identification were evaluated (see Chapter 4). It is clear, however, that the tools developed within this study can lead to a methodology for independent evaluation of search effectiveness. As stated earlier we feel that the approach employed is particularly promising now that adequate data reduction/analysis tools are in place.

5.5 ADDITIONAL RESULTS AND FINDINGS

This section is organized to summarize specific findings drawn from the data analysis of Chapter 4. Each finding is referenced back to the specific section of that chapter from which it was drawn.

- The logon process is still difficult for users and problems occur daily in this area (4.1.1, 4.1.2).
- More active users are associated with institutions having earlier start dates on the system (4.1.2).
- Institution type was (more) consistent among the most active users (4.1.2).
- Over 70% of all errors (corrected) are made in using single term, Boolean or stringsearching and display commands (4.2.1).
- Over 75% of "no posting" search results are accounted for by Boolean or single term searching (4.2.2).
- Analysis of longer sequences of activities leading to "no posting" results provided a large number of infrequently occurring sequences (4.2.2). See also findings of Section 4.3 and 4.4.2.1.
- Searchers in all three categories of experience use the display command at least part of the time to refine searches (4.2.3). See Wanger et al. (1980) for conflicting findings.

- The "80/20" rule applies in analyzing sequences of pairs of activities in that 80% of all occurrences are accounted for by about 20% of the pairs occurring (4.3).

- The level of specificity of activity codes must be applied selectively or a large number of infrequent activities (and sequences of activities) will occur (4.3). See also 4.2.2.

- The most frequently occurring sequences of activities are quite similar across user types for longer sequences (4.3).

- Stored search, dictionary look-up, and off-line search activities appear too infrequently to serve as distinguishing characteristics for the subsamples drawn (4.3).

- Term, advanced term, Boolean and display commands showed interesting distinguishing patterns between three classes of users (4.3 - Figure 4.9).

- First-order data provided significant differences between user groups for sequences starting with selected activities (4.3).

- Novices search more slowly (FENICHEL): agrees with our findings (4.4.1.1).

- Novices make more errors (FENICHEL): doesn't agree with our findings, but error definitions were different (4.4.1.2).

- Experienced searchers use more commands and connect time (FENICHEL): agrees with our findings for frequent and infrequent users but for moderate users our results don't agree (4.4.1.3).

- Experienced users employ relatively simple search strategies (FENICHEL): was not confirmed in our findings (4.4.1.5). See also CHAPMAN for agreement with our findings (4.4.2.3).

- Higher order analysis increases the number of unique fragments generated (CHAPMAN): was consistent with our findings (4.4.2.1). See also 4.2.2 and 4.3.

- Differences in low order data are sufficient to account for differences in higher order data (CHAPMAN): was not confirmed (4.4.2.2).

- Experienced searchers use more advanced search techniques (CHAPMAN): was in agreement with our findings (4.4.2.3) See also 4.4.1.5.

- Frequently occurring long sequences of identical commands for inexperienced users involve the display command (CHAPMAN): was in agreement with our findings (4.4.2.4).

- For experienced users the most frequently occurring long sequence of commands involved the advanced select feature (CHAPMAN): was not confirmed (4.4.2.4).

6.0 CONCLUSION

This study addressed a wide number of objectives and was not able to accomplish all of them. Despite this, a significant added understanding of how users perform on-line searches was gained. The tools developed to analyze audit-trail data are a major research resource and the ability to analyze such data through further stages of refinement holds great promise.

Of immediate interest to NLM should be the capability to capture and analyze audit-trail data, once parsed in a fashion that is both meaningful and yet protects the sensitivity of the raw data (see PENNIMAN, 1977, for further discussion of this issue). The data reduction techniques could be applied to all or part of NLM's audit trail data so that samples could be preserved for further research. This would eliminate the concern about distributing more widely the raw data and yet not limit further research in this area.

More research is needed in the area of search effectiveness measurement. While this study established some good baseline data, it did not provide the necessary results to build a general model of ineffective user behavior. Many of the characteristics of use leading to such a model were, however, identified. One area suggested for further evaluation is the variation in search patterns as a function of document set size retrieved. This data is available in the parsed records and could be evaluated. Another area of promise is the consolidation of activity sequences into "strings" following the work of CHAPMAN (1981). This consolidation would provide insight into how sequences of activities are concentrated. Finally, the data presented in Figure 4.7 representing frequently occurring patterns should be evaluated further for patterns unique to a single sample of users.

The research has also shown the value of this form of process analysis in that results of research on different search systems were able to be compared to identify invariant factors across systems with respect to searcher behavior.

Future research should place major emphasis on measures of search effectiveness beyond the traditional precision and recall parameters. A continuing research program at OCLC is addressing this as part of the broader area of "subject access." It is our hope that the techniques described in this report will provide a foundation for research projects in several institutions so that we can all benefit from the comparative findings.

REFERENCES CITED

Andreae, John H. 1977. Thinking with the Teachable Machine. New York, NY: Academic Press, Inc.; 1977. 178p. ISBN: 0-12-060050-1.

Brown, R.N.; Agrawala, A.K. 1974. On the Behavior of Users of the MEDLINE System. In: Changing Patterns in Information Retrieval. Washington, D.C.: American Society for Information Science; 1974: pp. 36-38.

Burton, Richard R.; Brown, John S. 1979. An Investigation of Computer Coaching for Informal Learning Activities. International Journal of Man-Machine Studies. 1979 January; 11(1): 5-24. ISSN: 0020-7373.

Chapman, Janet L. 1981. A State-Transition Analysis of Online Information Seeking Behavior. Journal of the American Society of Information Science. 1981 September; 32(5): 325-333. ISSN: 0002-8231.

Deo, Narsingh. 1974. Graph Theory with Applications to Engineering and Computer Science. Englewood Cliffs, NJ: Prentice Hall, Inc.; 1974. 478p. ISBN: 0-13-363473-6.

Dominick, W.D.; Penniman, W.D. 1979. Interaction Monitoring Considerations Within Network-Based Information Systems. Collected Papers Presented at the 9th ASIS Mid-Year Meeting, Pittsburgh, PA.; 1979 May 14-17.

Fenichel, Carol Hansen. 1981. On-Line Searching: Measures that Discriminate Among Users with Different Types of Experiences. Journal of the American Society of Information Science. 1981 January; 32(1): 23-32. ISSN: 0002-8231.

Gonzalez, Rafael C.; Thomason, Michael G. 1978. Syntactic Pattern Recognition. Reading, MA: Addison-Wesley Publishing Co.; 1978. 283p. ISBN: 0-201-02930-8.

Mann, William C. 1977. Man-Machine Communication Research: Final Report. 1977 February. Available from: Advanced Research Projects Agency, 1400 Wilson Blvd., Arlington, VA, 22209. ARPA Report No. ISI/RR-77-57.

Meadow, C., et al. 1977. Individualized Instruction for Data Access (IIDA). Philadelphia, PA: Drexel University: 1977 July; Final Design Report under NSF grant DSI-76-09737.

Penniman, W.D. 1975. A Stochastic Process Analysis of On-Line User Behavior. Proceedings of the American Society for Information Science Annual Meeting. 12: 1975.

Penniman, W.D. 1977. Suggestions for Systematic Evaluation of On-Line Monitoring Issues. Proceedings of the American Society for Information Science Annual Meeting. 14: 1977.

Penniman, W.D.; Dominick, W.D. 1979. The Application of Monitor Data Analysis Techniques to Audit Trail Data from an On-Line System. Collected Papers Presented at the 9th ASIS Mid-Year Meeting, Pittsburg, PA; 1979 May 14-17.

Penniman, W.D.; Dominick, W.D. 1980. Monitoring and Evaluation of On-Line Information System Usage. Information Processing and Management. 1980 January; 16(1): 17-35.

Penniman, W.D.; Perry, J.C. 1976. Tempo of On-Line User Interaction. Presented at 5th ASIS Mid-Year Meeting; 1976 May; Nashville, TN.

Relles, Nathan; Sondheimer, Norman K. 1981. Recent Advances in User Assistance. Paper presented at: Joint Conference on Easier and More Productive Use of Computing Systems; 1981 May 20; Ann Arbor, MI.

Rouse, William B. 1980. Systems Engineering Models of Human-Machine Interaction. New York, NY: Elsevier North Holland, Inc.; 1980. 152p. ISBN: 0-444-00366-5.

Stabell, Charles B. 1974. Individual Differences in Managerial Decision-Making Process: A study of conversational computer system usage. Boston, MA: Massachusetts Institute of Technology; 1974. Dissertation.

Wanger, Judith; Cuadra, Carlos A.; Fishburn, Mary. 1976. Impact of On-Line Retrieval Services: A Survey of Users, 1974-75. Santa Monica, CA: Systems Development Corporation; 1976. ISBN: 0-916368-01-7.

Wanger, Judith; McDonald, Dennis; Berger, Mary C. 1980. Evaluation of the On-Line Search Process: A Final Report. Final Report under contract to U.S. National Library of Medicine, National Institute of Health, of the Department of Health, Education and Welfare (Contract Number NO1-LM-8-4734).